Biometrika, *pp.* 1–6

Supplementary material for Inference in a Survival Cure Model with Mismeasured Covariates using a Simulation-Extrapolation Approach

BY AURELIE BERTRAND, CATHERINE LEGRAND

Institute of Statistics, Biostatistics and Actuarial Sciences, Université catholique de Louvain, Voie du Roman Pays 20, 1348 Louvain-la-Neuve, Belgium

aurelie.bertrand@uclouvain.be, catherine.legrand@uclouvain.be

RAYMOND J. CARROLL

Department of Statistics, Texas A&M University, College Station, 447 Blocker Building, Texas 77843-3143, U.S.A.

carroll@stat.tamu.edu

CHRISTOPHE DE MEESTER

Cardiovascular Research Group, Institute of Experimental and Clinical Research, Université catholique de Louvain, Avenue Hippocrate 55, 1200 Brussels, Belgium

christophe.demeester@uclouvain.be

AND INGRID VAN KEILEGOM

Institute of Statistics, Biostatistics and Actuarial Sciences, Université catholique de Louvain, Voie du Roman Pays 20, 1348 Louvain-la-Neuve, Belgium

ingrid.vankeilegom@uclouvain.be

1. PROOF OF THEOREM 1

For a fixed λ and a fixed b, we have from Theorem 1 in Zeng et al. (2006) that $\|\widehat{\beta}_{\lambda,b} - \beta_\lambda\| \to 0$, $\quad \sup_{t\in\mathbb{R}^+} |\widehat{F}_{\lambda,b}(t) - F_\lambda(t)| \to 0$, with probability 1. As a result, for a fixed λ, since $\widehat{\beta}_\lambda = B^{-1}\sum_{b=1}^B \widehat{\beta}_{\lambda,b}$ and $\widehat{F}_\lambda = B^{-1}\sum_{b=1}^B \widehat{F}_{\lambda,b}$, by Slutsky's theorem, $\|\widehat{\beta}_\lambda - \beta_\lambda\| \to 0$, $\sup_{t\in\mathbb{R}^+} |\widehat{F}_\lambda(t) - F_\lambda(t)| \to 0$ with probability 1.

An immediate result is that $\|\widehat{\beta}(\Lambda) - \beta(\Lambda)\| \to 0$, $\quad \sup_{t\in\mathbb{R}^+} |\widehat{F}(\Lambda,t) - F(\Lambda,t)| \to 0$, with probability 1, where $\beta(\Lambda) = (\beta_{\lambda_1}^{\mathrm{T}}, \ldots, \beta_{\lambda_K}^{\mathrm{T}})^{\mathrm{T}}$ and $F(\Lambda,t) = \{F_{\lambda_1}(t), \ldots, F_{\lambda_K}(t)\}^{\mathrm{T}}$, for a finite grid $\Lambda = (\lambda_1, \ldots, \lambda_K)^{\mathrm{T}}$.

We now take the extrapolation step of the simex algorithm into account. We first show the consistency of $\widehat{\beta}_{\mathrm{SIMEX}}$. Similarly to what is done in Li & Lin (2003), suppose that β_λ can be specified using a parametric model $g_\beta(\gamma_\beta, \lambda)$ depending on a vector of parameters γ_β. Under the assumption that this is the true extrapolation function, we have that $\beta_{\mathrm{TRUE}} = g_\beta(\gamma_\beta, -1)$ and $\widehat{\beta}_{\mathrm{SIMEX}} = g_\beta(\widehat{\gamma}_\beta, -1)$, where $\widehat{\gamma}_\beta$ solves, by the least squares estimation method, $\dot{g}_\beta(\gamma_\beta, \Lambda)^{\mathrm{T}}\{g_\beta(\gamma_\beta, \Lambda) - \widehat{\beta}(\Lambda)\} = 0$ and $\dot{g}_\beta(\gamma_\beta, \Lambda)$ is the $PK \times \dim(\gamma_\beta)$ matrix of partial derivatives of the elements of $g_\beta(\gamma_\beta, \Lambda)$ with respect to the elements of γ_β.

We then have that

$$\widehat{\gamma}_\beta - \gamma_\beta = \left\{\dot{g}_\beta(\gamma_\beta,\Lambda)^{\mathrm{T}}\dot{g}_\beta(\gamma_\beta,\Lambda)\right\}^{-1}\dot{g}_\beta(\gamma_\beta,\Lambda)^{\mathrm{T}}\{\widehat{\beta}(\Lambda)-\beta(\Lambda)\} + o_p(1)$$

and hence, with probability 1, $\|\widehat{\gamma}_\beta - \gamma_\beta\| \to 0$, $n \to \infty$, by the continuous mapping theorem, if $\dot{g}_\beta(\gamma_\beta,\Lambda)$ is bounded and $\dot{g}_\beta(\gamma_\beta,\Lambda)^{\mathrm{T}}\dot{g}_\beta(\gamma_\beta,\Lambda)$ is invertible. Finally, by using once again the continuous mapping theorem, since $\widehat{\beta}_{\text{SIMEX}} = g_\beta(\widehat{\gamma}_\beta,-1)$ and $\beta_{\text{TRUE}} = g_\beta(\gamma_\beta,-1)$, we find that, with probability 1,

$$\|\widehat{\beta}_{\text{SIMEX}} - \beta_{\text{TRUE}}\| \to 0.$$

We finally turn to the proof of the consistency of $\widehat{F}_{\text{SIMEX}}$. Similarly to what is done in Li & Lin (2003), suppose that, for a fixed t, $F_\lambda(t)$ can be specified using a parametric model $g_t(\gamma_t,\lambda)$ depending on a parameter vector γ_t. If this is the true extrapolation function, we have $F_{\text{TRUE}}(t) = g_t(\gamma_t,-1)$ and $\widehat{F}_{\text{SIMEX}}(t) = g_t(\widehat{\gamma}_t,-1)$, where $\widehat{\gamma}_t$ is a solution of $\dot{g}_t(\gamma_t,\Lambda)^{\mathrm{T}}\{g_t(\gamma_t,\Lambda) - \widehat{F}(\Lambda,t)\} = 0$ and $\dot{g}_t(\gamma_t,\Lambda) = \partial g_t(\gamma_t,\Lambda)/\partial\gamma_t^{\mathrm{T}}$. It now follows that

$$\widehat{\gamma}_t - \gamma_t = \left\{\dot{g}_t(\gamma_t,\Lambda)^{\mathrm{T}}\dot{g}_t(\gamma_t,\Lambda)\right\}^{-1}\dot{g}_t(\gamma_t,\Lambda)^{\mathrm{T}}\left\{\widehat{F}(\Lambda,t) - F(\Lambda,t)\right\} + o_p(1)$$

and hence, with probability 1, $\sup_{t\in\mathbb{R}^+}\|\widehat{\gamma}_t - \gamma_t\| \to 0$, by the continuous mapping theorem, assuming that $\dot{g}_t(\gamma_t,\Lambda)$ is bounded uniformly in t and that $\dot{g}_t(\gamma_t,\Lambda)^{\mathrm{T}}\dot{g}_t(\gamma_t,\Lambda)$ is invertible. Finally, by using once again the continuous mapping theorem, since $\widehat{F}_{\text{SIMEX}}(t) = g_t(\widehat{\gamma}_t,-1)$ and $F_{\text{TRUE}}(t) = g_t(\gamma_t,-1)$, we find that, with probability 1,

$$\sup_{t\in\mathbb{R}^+}|\widehat{F}_{\text{SIMEX}}(t) - F_{\text{TRUE}}(t)| \to 0.$$

2. SIMULATION RESULTS

Table 1 summarizes the results obtained for the 4 settings described in Section 4·2 with $\mu = 1.0$, while Table 2 pertains to the settings of Section 4·3 with $\mu = 1.0$.

Table 3 reports the results of a simulation study investigating the impact of the choice of the grid of values for λ in the simex method. We still consider the same setting as in Subsection 4·4, and compare the estimates obtained by using $\lambda \in \{0, 0.5, 1, 1.5, 2\}$, as in all previous simulation settings, with those obtained with $\lambda \in \{0, 0.25, 0.5, 0.75, 1, 1.25, 1.5, 1.75, 2\}$. The quadratic extrapolation function was used. It appears that both grids of λ values yield nearly identical results in this setting.

Table 1: Empirical bias, empirical and estimated variances, coverages and mean squared errors for the settings with one mismeasured covariate, when $\mu = 1.0$

			Ma & Yin method			Naive method			Simex method		
n	v	Estimate	β_0	β_1	β_2	β_0	β_1	β_2	β_0	β_1	β_2
200	0.1	Bias	0.2	1.8	−0.6	5.4	−10.2	−0.2	0.7	0.4	−0.5
		Emp. var.	4.1	10.2	3.0	3.6	7.7	3.0	4.1	9.8	3.0
		Est. var.	4.2	10.0	2.8	3.7	7.6	2.8	4.1	9.6	2.8
		95% cv	94.8	94.6	94.2	94.0	92.8	94.6	94.8	94.4	94.2
		MSE	4.1	10.2	3.0	3.9	8.7	3.0	4.1	9.8	3.0
200	0.2	Bias	−1.4	5.9	−0.8	15.1	−32.3	0.3	4.6	−8.5	−0.2
		Emp. var.	5.2	17.1	3.2	3.2	6.0	3.0	4.2	11.3	3.1
		Est. var.	5.6	16.7	3.0	3.2	5.7	2.8	4.1	9.8	2.8
		95% cv	96.4	96.0	94.0	87.0	73.0	94.8	92.8	91.6	94.0
		MSE	5.2	17.4	3.2	5.5	16.4	3.0	4.4	12.0	3.1
300	0.1	Bias	−0.4	2.4	−0.9	4.8	−9.4	−0.6	0.1	1.3	−0.9
		Emp. var.	3.1	7.1	2.0	2.8	5.4	2.0	3.0	6.9	2.0
		Est. var.	2.8	6.6	1.9	2.4	5.0	1.8	2.7	6.4	1.9
		95% cv	95.4	94.6	95.6	91.6	90.4	95.0	94.4	94.2	95.4
		MSE	3.1	7.1	2.1	3.0	6.3	2.0	3.0	6.9	2.1
300	0.2	Bias	−1.3	4.9	−1.2	14.6	−31.6	−0.1	3.9	−7.6	−0.9
		Emp. var.	3.7	10.4	2.2	2.4	3.9	2.0	3.1	7.5	2.1
		Est. var.	3.5	10.4	2.0	2.1	3.8	1.8	2.7	6.5	1.9
		95% cv	96.4	96.6	94.8	82.5	63.8	94.0	92.2	92.4	94.7
		MSE	3.7	10.6	2.2	4.5	13.9	2.0	3.3	8.1	2.1

Emp. var., empirical variance; Est. var., estimated variance; 95% cv, coverage probabilities of 95% confidence intervals computed based on the asymptotic normal distribution; MSE, mean squared error. NOTE: All numbers were multiplied by 100.

Table 2: Empirical bias, empirical and estimated variances, coverages and mean squared errors for the settings with two mismeasured covariates, when $\mu = 1.0$

	Ma & Yin method				Naive method				Simex method			
Estimate	β_0	β_1	β_2	β_3	β_0	β_1	β_2	β_3	β_0	β_1	β_2	β_3
					$n = 200$, $v_1 = 0.1$, $v_2 = 0.1$							
Bias	0.4	0.7	1.3	−0.6	5.1	−11.7	−0.9	−0.1	1.1	−1.1	1.0	−0.5
Emp. var.	5.7	11.8	1.3	3.2	5.0	8.8	1.2	3.1	5.6	11.3	1.3	3.1
Est. var.	4.9	10.9	1.2	3.1	4.2	8.1	1.1	3.0	4.7	10.3	1.2	3.1
95% cv	93.0	94.6	94.8	95.8	91.8	92.6	94.8	96.0	92.4	94.8	94.2	96.0
MSE	5.7	11.8	1.3	3.2	5.3	10.1	1.2	3.1	5.6	11.3	1.3	3.1
					$n = 200$, $v_1 = 0.2$, $v_2 = 0.2$							
Bias	−0.9	5.0	2.4	−1.1	13.5	−34.6	−6.0	1.0	5.2	−11.1	0.3	−0.3
Emp. var.	7.6	20.5	1.7	3.5	4.6	6.8	1.1	3.1	6.2	13.4	1.4	3.3
Est. var.	6.6	18.7	1.5	3.4	3.7	6.1	1.0	3.0	4.8	10.7	1.3	3.2
95% cv	94.4	95.2	95.0	95.2	86.4	71.0	90.2	95.2	91.0	90.8	95.2	94.8
MSE	7.6	20.8	1.7	3.6	6.4	18.7	1.5	3.1	6.4	14.6	1.4	3.3
					$n = 300$, $v_1 = 0.1$, $v_2 = 0.1$							
Bias	1.6	1.2	0.9	−0.9	6.2	−11.0	−1.3	−0.3	2.0	−0.1	0.7	−0.8
Emp. var.	3.2	7.6	0.8	2.0	2.8	5.8	0.7	1.9	3.2	7.5	0.8	2.0
Est. var.	3.2	7.1	0.8	2.0	2.8	5.4	0.7	2.0	3.1	6.8	0.8	2.0
95% cv	95.8	94.0	96.0	95.6	93.2	92.8	95.2	95.4	94.8	94.0	96.0	95.8
MSE	3.2	7.6	0.8	2.0	3.2	7.0	0.7	1.9	3.2	7.5	0.8	2.0
					$n = 300$, $v_1 = 0.2$, $v_2 = 0.2$							
Bias	0.5	4.5	1.7	−1.1	14.6	−33.8	−6.3	0.7	6.2	−9.7	0.2	−0.6
Emp. var.	4.2	12.5	1.0	2.1	2.5	4.5	0.7	1.9	3.5	8.7	0.9	2.1
Est. var.	4.2	11.7	1.0	2.2	2.5	4.0	0.7	2.0	3.2	7.1	0.9	2.1
95% cv	95.4	94.8	96.4	95.4	84.8	60.8	87.2	95.4	92.0	90.6	96.0	95.0
MSE	4.2	12.7	1.0	2.1	4.6	15.9	1.1	1.9	3.9	9.6	0.9	2.1

Emp. var., empirical variance; Est. var., estimated variance; 95% cv, coverage probabilities of 95% confidence intervals computed based on the asymptotic normal distribution; MSE, mean squared error.
NOTE: All numbers were multiplied by 100.

Table 3: Empirical bias, empirical and estimated variances, coverages and mean squared errors for the simulations investigating the robustness of simex with respect to the grid of λ

			$\lambda \in \{0, 0.5, 1, 1.5, 2\}$			$\lambda \in \{0, 0.25, 0.5, 0.75, 1, 1.25, 1.5, 1.75, 2\}$		
n	v		β_0	β_1	β_2	β_0	β_1	β_2
200	0.1	Bias	−1.4	−0.3	−2.2	−1.4	−0.3	−2.2
		Emp. var.	12.4	20.6	6.7	12.4	20.6	6.7
		Est. Var.	10.8	19.1	6.3	10.7	19.1	6.3
		95% cv	93.4	95.8	94.8	93.4	96.2	94.8
		MSE	12.4	20.6	6.8	12.4	20.6	6.8
200	0.25	Bias	7.2	−18.2	−2.0	7.4	−18.6	−2.0
		Emp. var.	12.1	21.0	6.8	12.1	20.7	6.8
		Est. Var.	9.7	14.8	6.3	9.7	14.6	6.3
		95% cv	92.7	88.8	95.1	92.7	88.6	95.1
		MSE	12.7	24.3	6.8	12.6	24.2	6.8
300	0.1	Bias	−2.5	1.3	−1.3	−2.5	1.3	−1.3
		Emp. var.	9.9	16.5	4.3	9.8	16.4	4.3
		Est. Var.	7.6	12.5	4.1	7.6	12.5	4.1
		95% cv	92.8	94.0	95.6	92.8	93.8	95.6
		MSE	10.0	16.5	4.4	9.9	16.4	4.4
300	0.25	Bias	5.8	−15.3	−1.4	6.1	−15.6	−1.5
		Emp. var.	10.0	17.7	4.4	9.9	17.4	4.4
		Est. Var.	7.0	9.8	4.2	6.9	9.7	4.2
		95% cv	89.7	84.3	95.6	89.7	83.9	95.6
		MSE	10.3	20.0	4.5	10.3	19.9	4.4

Emp. var., empirical variance; Est. var., estimated variance; 95% cv, coverage probabilities of 95% confidence intervals computed based on the asymptotic normal distribution; MSE, mean squared error. NOTE: All numbers were multiplied by 100.

References

Li, Y. & Lin, X. (2003). Functional inference in frailty measurement error models for clustered survival data using the simex approach. *J. Am. Statist. Assoc.* **98**, 191–203.

Zeng, D., Yin, G. & Ibrahim, J. G. (2006). Semiparametric transformation models for survival data with a cure fraction. *J. Am. Statist. Assoc.* **101**, 670–684.