Web-based Supplemental Materials for *Simultaneous Variable Selection for Joint Models of Longitudinal and Survival Outcomes*

by

Zangdong He, Wanzhu Tu, Sijian Wang, Haoda Fu and Zhangsheng Yu

1 Expectation conditional maximization procedures to optimize the penalized likelihood

Let $\mathbf{\Theta} = (\boldsymbol{\theta}, \zeta_{1m}, \eta_{2l})$, where $\boldsymbol{\theta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \mathbf{\Gamma}_1, \mathbf{\Gamma}_2, \boldsymbol{\phi})$ are defined in section 2.2. The expectation conditional maximization procedures to optimize the penalized likelihood are proposed as follows:

1. Initialize $(\boldsymbol{\beta_1^{(0)}}, \boldsymbol{\beta_2^{(0)}}, \boldsymbol{\gamma_{1m}^{(0)}}, \zeta_{1m}^{(0)}, \boldsymbol{\gamma_{2l}^{(0)}}, \eta_{2l}^{(0)}, \boldsymbol{\phi^{(0)}})$ with some plausible values.
2. For iteration s, update $\boldsymbol{\beta_1}, \boldsymbol{\beta_2}$ by adaptive LASSO,

$$\begin{aligned}\boldsymbol{\beta_1}^{(s)}, \boldsymbol{\beta_2}^{(s)} = &\underset{\boldsymbol{\beta_1},\boldsymbol{\beta_2}}{argmax}\, \tilde{Q}(\boldsymbol{\beta_1}, \boldsymbol{\beta_2}, \boldsymbol{\hat{\Gamma}_1^{(s-1)}}, \boldsymbol{\hat{\Gamma}_2^{(s-1)}}, \boldsymbol{\hat{\phi}^{(s-1)}} | \boldsymbol{\hat{\beta}_1^{(s-1)}}, \boldsymbol{\hat{\beta}_2^{(s-1)}}, \boldsymbol{\hat{\Gamma}_1^{(s-1)}}, \boldsymbol{\hat{\Gamma}_2^{(s-1)}}, \boldsymbol{\hat{\phi}^{(s-1)}}) \\ &- \lambda_1 \sum_{j=1}^{p} \omega_{\beta_{1j}} |\beta_{1j}| - \lambda_2 \sum_{k=1}^{p} \omega_{\beta_{2k}} |\beta_{2k}|.\end{aligned}$$

3. update $\boldsymbol{\gamma_{1m}}, \boldsymbol{\gamma_{2l}}$:

$$\begin{aligned}\boldsymbol{\gamma_{1m}}^{(s)}, \boldsymbol{\gamma_{2l}}^{(s)} = &\underset{\gamma_{1m},\gamma_{2l}}{argmax}\, \tilde{Q}(\boldsymbol{\hat{\beta}_1^{(s)}}, \boldsymbol{\hat{\beta}_2^{(s)}}, \boldsymbol{\Gamma_1}, \boldsymbol{\Gamma_2}, \boldsymbol{\hat{\phi}^{(s-1)}} | \boldsymbol{\hat{\beta}_1^{(s)}}, \boldsymbol{\hat{\beta}_2^{(s)}}, \boldsymbol{\hat{\Gamma}_1^{(s-1)}}, \boldsymbol{\hat{\Gamma}_2^{(s-1)}}, \boldsymbol{\hat{\phi}^{(s-1)}}) \\ &- \frac{1}{4} \sum_{m=2}^{q} \frac{(\lambda_3 \omega_{\boldsymbol{\gamma_{1m}}})^2}{(\zeta_{1m}^{(s-1)})^2} ||\boldsymbol{\gamma_{1m}}||^2 - \frac{1}{4} \sum_{l=2}^{q} \frac{(\lambda_4 \omega_{\boldsymbol{\gamma_{2l}}})^2}{(\eta_{2l}^{(s-1)})^2} ||\boldsymbol{\gamma_{2l}}||^2.\end{aligned}$$

4. update ζ_{1m}, η_{2l}:

$$\zeta_{1m}^{(s)} = \sqrt{\frac{\lambda_{\gamma_1}\omega_{\gamma_{1m}}}{2}||\boldsymbol{\gamma_{1m}^{(s)}}||}, \eta_{2l}^{(s)} = \sqrt{\frac{\lambda_{\gamma_2}\omega_{\gamma_{2l}}}{2}||\boldsymbol{\gamma_{2l}^{(s)}}||}.$$

5. update $\boldsymbol{\phi}$:

$$\boldsymbol{\phi} = \underset{\boldsymbol{\phi}}{argmax}\, \tilde{Q}(\boldsymbol{\hat{\beta}_1^{(s)}, \hat{\beta}_2^{(s)}, \hat{\Gamma}_1^{(s)}, \hat{\Gamma}_2^{(s)}}, \boldsymbol{\phi}|\boldsymbol{\hat{\beta}_1^{(s)}, \hat{\beta}_2^{(s)}, \hat{\Gamma}_1^{(s)}, \hat{\Gamma}_2^{(s)}}, \boldsymbol{\hat{\phi}^{(s-1)}}).$$

6. Terminate the iteration when $max|\boldsymbol{\Theta}^{(s)} - \boldsymbol{\Theta}^{(s-1)}|$ are small enough. Otherwise, let $s = s + 1$ and go back to step 2.

Before updating parameters in each step, the corresponding $\tilde{Q}$ function is approximated by Gaussian quadrature in the E-step. To improve computation stability, smaller subset of $(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\Gamma}_1, \boldsymbol{\Gamma}_2, \boldsymbol{\phi})$ could be updated iteratively. We could update $\boldsymbol{\beta}_1$ when $(\boldsymbol{\beta}_2, \boldsymbol{\Gamma}_1, \boldsymbol{\Gamma}_2, \boldsymbol{\phi})$ is fixed, and then update $\boldsymbol{\beta}_2$ when $(\boldsymbol{\beta}_1, \boldsymbol{\Gamma}_1, \boldsymbol{\Gamma}_2, \boldsymbol{\phi})$ is fixed, and sequentially for $\boldsymbol{\Gamma}_1$, $\boldsymbol{\Gamma}_2$, and $\boldsymbol{\phi}$ when other parameters are fixed. It is at the price of more iterations.

2 Data generation for simulation study: Scenario 5

In Scenario 5, we generate the longitudinal outcome Y_{ij} from the following model:

$$\begin{aligned} Y_{ij} =& 1 + 1.5X_{1ij,1} + 2X_{1ij,2} + 0X_{1ij,3} + 0X_{1ij,4} + b_{li,0} \\ &+ b_{li,1}Z_{1ij,1} + b_{li,2}Z_{1ij,2} + b_{li,3}Z_{1ij,3} + b_{li,4}Z_{1ij,4} + \epsilon_{ij}, \end{aligned}$$

and the failure time from a Weibull distribution with the hazard function:

$$\begin{aligned} \lambda_i(t) = \lambda_0(t) \exp(&1.5x_{2i,1} + 2x_{2i,2} + 0x_{2i,3} + 0x_{2i,4} \\ &+ b_{si,0} + b_{si,1}z_{2i,1} + b_{si,2}z_{2i,2} + b_{si,3}z_{2i,3} + b_{si,4}z_{2i,4}), \end{aligned}$$

for $i = 1, \ldots, 800, j = 1, \ldots, 5$, where $\lambda_0(t) = \alpha\lambda t^{\alpha-1}$ with $\alpha = 2$, and $\lambda = \exp(1) = 2.718$.

Random effect $\boldsymbol{b_i}$ is independently generated from $N(0, \boldsymbol{I_5})$. $\boldsymbol{b_{li}} = (b_{li,0}, b_{li,1}, b_{li,2}, b_{li,3}, b_{li,4})$ is obtained by $\boldsymbol{b_{li} = \Gamma_1 b_i}$ and $\boldsymbol{b_{si}} = (b_{si,0}, b_{si,1}, b_{si,2}, b_{si,3}, b_{si,4})$ is obtained by $\boldsymbol{b_{si}} = \boldsymbol{\Gamma_2 b_i}$, where

$$\boldsymbol{\Gamma_1} = \boldsymbol{\Gamma_2} = \sigma_D \begin{Bmatrix} 1 & 0 & 0 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{Bmatrix}^{\frac{1}{2}}$$

and $\sigma_D = \sqrt{0.5}$. Covariates $X_{1ij,1} = Z_{1ij,1}, X_{1ij,2} = Z_{1ij,2}, X_{1ij,3} = Z_{1ij,3}, X_{1ij,4} = Z_{1ij,4}$ and $x_{2i,1} = z_{2i,1}$, $x_{2i,2} = z_{2i,2}, x_{2i,3} = z_{2i,3}, x_{2i,4} = z_{2i,4}$ are generated as independent $N(0,1)$ variables; The measurement error $\epsilon_{ij} \sim i.i.d.N(0,1)$. The censoring time is independently generated from an exponential distribution to achieve a 60% censoring percentage.

3 Data generation for simulation study: Scenario 6

In Scenario 6, we generate the longitudinal outcome Y_{ij} from the following model:

$$\begin{aligned} Y_{ij} =& 1 + 1.5X_{1ij,1} + 2X_{1ij,2} + 2.5X_{1ij,3} + 0X_{1ij,4} + 0X_{1ij,5} + 0X_{1ij,6} + 0X_{1ij,7} + \\ & b_{li,0} + b_{li,1}Z_{1ij,1} + b_{li,2}Z_{1ij,2} + b_{li,3}Z_{1ij,3} + b_{li,4}Z_{1ij,4} + b_{li,5}Z_{1ij,5} + b_{li,6}Z_{1ij,6} + \\ & b_{li,7}Z_{1ij,7} + \epsilon_{ij}, \end{aligned}$$

and the failure time from a Weibull distribution with the hazard function:

$$\begin{aligned} \lambda_i(t) = \lambda_0(t) \exp(& 1.5x_{2i,1} + 2x_{2i,2} + 2.5x_{2i,3} + 0x_{2i,4} + 0x_{2i,5} + 0x_{2i,6} + 0x_{2i,7} + \\ & b_{si,0} + b_{si,1}z_{2i,1} + b_{si,2}z_{2i,2} + b_{si,3}z_{2i,3} + b_{si,4}z_{2i,4} + b_{si,5}z_{2i,5} + \\ & b_{si,6}z_{2i,6} + b_{si,7}z_{2i,7}), \end{aligned}$$

for $i = 1, \ldots, 250, j = 1, \ldots, 5$, where $\lambda_0(t) = \alpha\lambda t^{\alpha-1}$ with $\alpha = 2$, and $\lambda = \exp(1) = 2.718$.

Random effect $\boldsymbol{b_i}$ is independently generated from $N(0, \boldsymbol{I_8})$. $\boldsymbol{b_{li}} = (b_{li,0}, b_{li,1}, b_{li,2}, b_{li,3}, b_{li,4}\ b_{li,5}, b_{li,6}, b_{li,7})$ is obtained by $\boldsymbol{b_{li}} = \boldsymbol{\Gamma_1 b_i}$ and $\boldsymbol{b_{si}} = (b_{si,0}, b_{si,1}, b_{si,2}, b_{si,3}, b_{si,4},$

$b_{si,5}, b_{si,6}, b_{si,7})$ is obtained by $\boldsymbol{b_{si}} = \boldsymbol{\Gamma_2 b_i}$, where

$$\boldsymbol{\Gamma_1} = \boldsymbol{\Gamma_2} = \sigma_D \left\{ \begin{array}{cccccccc} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array} \right\}^{\frac{1}{2}}$$

and $\sigma_D = \sqrt{0.5}$. Covariates $X_{1ij,1} = Z_{1ij,1}, X_{1ij,2} = Z_{1ij,2}, X_{1ij,3} = Z_{1ij,3}, X_{1ij,4} = Z_{1ij,4}, X_{1ij,5} = Z_{1ij,5}$, $X_{1ij,6} = Z_{1ij,6}, X_{1ij,7} = Z_{1ij,7}$ and $x_{2i,1} = z_{2i,1}$, $x_{2i,2} = z_{2i,2}, x_{2i,3} = z_{2i,3}, x_{2i,4} = z_{2i,4}, x_{2i,5} = z_{2i,5}$, $x_{2i,6} = z_{2i,6}, x_{2i,7} = z_{2i,7}$ are generated as independent $N(0,1)$ variables; The measurement error $\epsilon_{ij} \sim i.i.d.N(0,1)$. The censoring time is independently generated from an exponential distribution to ahieve a 30% censoring percentage.

Web Table 1: Selection frequency of mixed effects in longitudinal and survival components for Scenario 5

Fixed effect selection							
Sel. Freq.(%) for Longitudinal component				Sel. Freq.(%) for Survival component			
$X_{1,1}$	$X_{1,2}$	$X_{1,3}$	$X_{1,4}$	$X_{2,1}$	$X_{2,2}$	$X_{2,3}$	$X_{2,4}$
Non-Zero	Non-Zero	Zero	Zero	Non-Zero	Non-Zero	Zero	Zero
100	100	0	0	100	100	0	0

Random effect selection							
Sel. Freq.(%) for Longitudinal component				Sel. Freq.(%) for Survival component			
$Z_{1,1}$	$Z_{1,2}$	$Z_{1,3}$	$Z_{1,4}$	$Z_{2,1}$	$Z_{2,2}$	$Z_{2,3}$	$Z_{2,4}$
Non-Zero	Non-Zero	Zero	Zero	Non-Zero	Non-Zero	Zero	Zero
100	100	0	0	99	99	1	0

Web Table 2: Estimation of fixed effects $\beta_{1,j}$ and $\beta_{2,j}$ in longitudinal and survival components for Scenario 5

	$\hat{\beta}_{1,j} \pm SE$(Coverage probability) for Longitudinal component[a]				
	Intercept	$X_{1,1}$	$X_{1,2}$	$X_{1,3}$	$X_{1,4}$
True value β	1	1.5	2	0	0
W/O selection $\hat{\beta}$	0.995±0.033(94%)	1.500±0.036(95%)	1.998±0.039(94%)	0.002	0.001
1^{st} stage $\hat{\beta}$	0.993±0.033(92%)	1.487±0.036(90%)	1.986±0.039(87%)	0.000	0.000
2^{nd} stage $\hat{\beta}$	0.999±0.029(95%)	1.505±0.034(95%)	2.001±0.036(91%)	0.000	0.000

	$\hat{\beta}_{2,j} \pm SE$(Coverage probability) for Survival component[a]				
	Intercept	$X_{1,1}$	$X_{1,2}$	$X_{1,3}$	$X_{1,4}$
True value β	-	1.5	2	0	0
W/O selection $\hat{\beta}$		1.355±0.126(74%)	1.844±0.149(75%)	0.008	0.019
1^{st} stage $\hat{\beta}$		0.989±0.125(0%)	1.381±0.145(1%)	0.000	0.000
2^{nd} stage $\hat{\beta}$		1.348±0.130(67%)	1.823±0.152(73%)	0.000	0.000

[a] $\hat{\beta}s$ are the averages of estimates over the 100 data sets; SE is the empirical standard error of the 100 $\hat{\beta}s$; For each data set, the 95% confidence interval based on the parameter and standard error estimates is calculated and the corresponding coverage probabilities for the true value over the 100 data sets are included in the parentheses. SE and coverage probability are only reported for non-zero variables.

Web Table 3: Estimation of random effects $\sqrt{D_{1kk}}$ and $\sqrt{D_{2kk}}$ in longitudinal and survival components for Scenario 5

	$\sqrt{\hat{D}_{1kk}}$ for Longitudinal component[a]					$\sqrt{\hat{D}_{2kk}}$ for Survival component[a]				
	$Intercept_1$	$Z_{1,1}$	$Z_{1,2}$	$Z_{1,3}$	$Z_{1,4}$	$Intercept_2$	$Z_{2,1}$	$Z_{2,2}$	$Z_{2,3}$	$Z_{2,4}$
True value $\sqrt{D_{kk}}$	0.707	0.707	0.707	0	0	0.707	0.707	0.707	0	0
W/O selection $\sqrt{\hat{D}_{kk}}$	0.791	0.817	0.817	0.052	0.050	0.776	0.820	0.825	0.205	0.202
1^{st} stage $\sqrt{\hat{D}_{kk}}$	0.787	0.773	0.763	0.000	0.000	0.407	0.368	0.319	0.000	0.000
2^{nd} stage $\sqrt{\hat{D}_{kk}}$	0.682	0.692	0.696	0.000	0.000	0.638	0.665	0.674	0.004	0.000

[a] $\sqrt{\hat{D}_{1kk}}$ and $\sqrt{\hat{D}_{2kk}}$ are the averages of estimates over the 100 data sets.

Web Table 4: Selection frequency of mixed effects in longitudinal and survival components for Scenario 6

Fixed effect selection													
Sel. Freq.(%) for Longitudinal component							Sel. Freq.(%) for Survival component						
$X_{1,1}$ Non-Zero	$X_{1,2}$ Non-Zero	$X_{1,3}$ Non-Zero	$X_{1,4}$ Zero	$X_{1,5}$ Zero	$X_{1,6}$ Zero	$X_{1,7}$ Zero	$X_{2,1}$ Non-Zero	$X_{2,2}$ Non-Zero	$X_{2,3}$ Non-Zero	$X_{2,4}$ Zero	$X_{2,5}$ Zero	$X_{2,6}$ Zero	$X_{2,7}$ Zero
100	100	100	0	0	0	0	100	100	100	0	0	0	0

Random effect selection													
Sel. Freq.(%) for Longitudinal component							Sel. Freq.(%) for Survival component						
$Z_{1,1}$ Non-Zero	$Z_{1,2}$ Non-Zero	$Z_{1,3}$ Non-Zero	$Z_{1,4}$ Zero	$Z_{1,5}$ Zero	$Z_{1,6}$ Zero	$Z_{1,7}$ Zero	$Z_{2,1}$ Non-Zero	$Z_{2,2}$ Non-Zero	$Z_{2,3}$ Non-Zero	$Z_{2,4}$ Zero	$Z_{2,5}$ Zero	$Z_{2,6}$ Zero	$Z_{2,7}$ Zero
100	100	100	0	0	0	0	97	93	94	6	4	1	9

Web Table 5: Estimation of fixed effects $\beta_{1,j}$ and $\beta_{2,j}$ in longitudinal and survival components for Scenario 6

	$\hat{\beta}_{1,j} \pm SE$(Coverage probability) for Longitudinal component[a]							
	Intercept	$X_{1,1}$	$X_{1,2}$	$X_{1,3}$	$X_{1,4}$	$X_{1,5}$	$X_{1,6}$	$X_{1,7}$
True value β	1	1.5	2	2.5	0	0	0	0
W/O selection $\hat{\beta}$	0.994±0.068(85%)	1.498±0.081(75%)	1.999±0.072(79%)	2.496±0.072(81%)	0.001	-0.004	0.000	-0.003
1^{st} stage $\hat{\beta}$	0.987±0.068(89%)	1.454±0.079(82%)	1.960±0.072(87%)	2.462±0.072(87%)	0.000	0.000	0.000	0.000
2^{nd} stage $\hat{\beta}$	0.994±0.064(87%)	1.497±0.076(82%)	1.995±0.074(85%)	2.496±0.073(86%)	0.000	0.000	0.000	0.000

	$\hat{\beta}_{2,j} \pm SE$(Coverage probability) for Survival component[a]						
	$X_{2,1}$	$X_{2,2}$	$X_{2,3}$	$X_{2,4}$	$X_{2,5}$	$X_{2,6}$	$X_{2,7}$
True value β	1.5	2	2.5	0	0	0	0
W/O selection $\hat{\beta}$	1.966±0.286(63%)	2.667±0.377(49%)	3.313±0.429(49%)	0.014	-0.025	0.011	0.035
1^{st} stage $\hat{\beta}$	1.039±0.249(20%)	1.495±0.331(28%)	1.897±0.370(30%)	0.000	0.000	0.000	0.000
2^{nd} stage $\hat{\beta}$	1.549±0.358(86%)	2.112±0.593(82%)	2.625±0.712(84%)	0.000	0.000	0.000	0.000

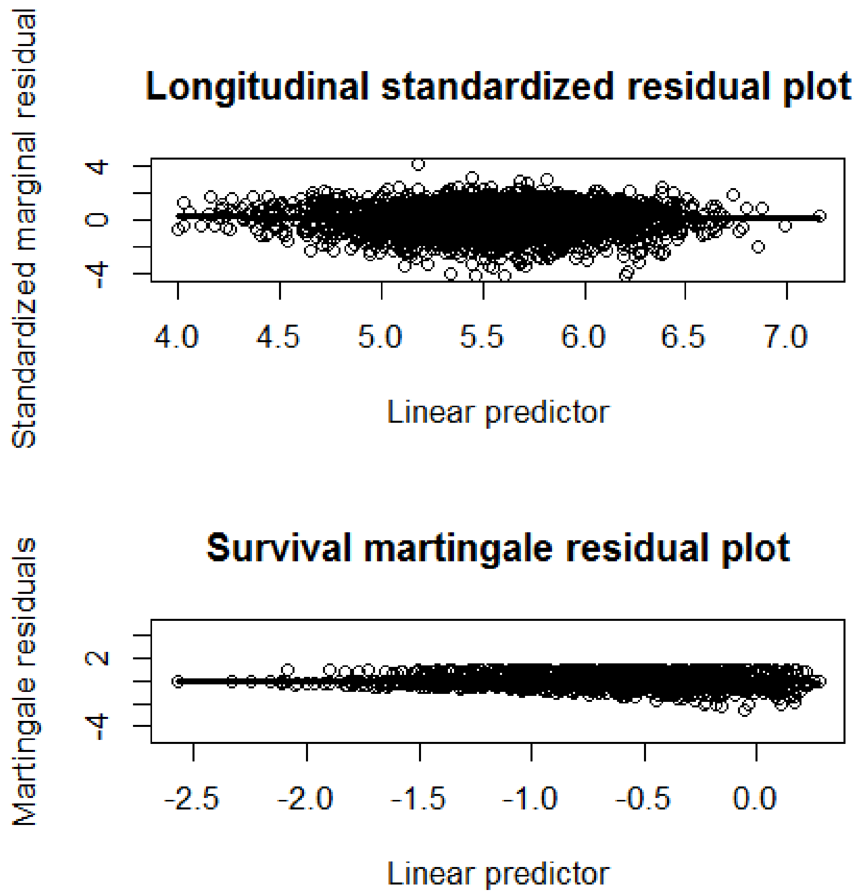
[a] $\hat{\beta}s$ are the averages of estimates over the 100 data sets; SE is the empirical standard error of the 100 $\hat{\beta}s$; For each data set, the 95% confidence interval based on the parameter and standard error estimates is calculated and the corresponding coverage probabilities for the true value over the 100 data sets are included in the parentheses. SE and coverage probability are only reported for non-zero variables.

Web Table 6: Estimation of random effects $\sqrt{D_{1kk}}$ and $\sqrt{D_{2kk}}$ in longitudinal and survival components for Scenario 6

	$\sqrt{\hat{D}_{1kk}}$ for Longitudinal component[a]							
	Intercept	$Z_{1,1}$	$Z_{1,2}$	$Z_{1,3}$	$Z_{1,4}$	$Z_{1,5}$	$Z_{1,6}$	$Z_{1,7}$
True value $\sqrt{D_{kk}}$	0.707	0.707	0.707	0.707	0	0	0	0
W/O selection $\sqrt{\hat{D}_{kk}}$	0.785	0.821	0.831	0.829	0.165	0.174	0.159	0.158
1^{st} stage $\sqrt{\hat{D}_{kk}}$	0.768	0.677	0.657	0.633	0.000	0.000	0.000	0.000
2^{nd} stage $\sqrt{\hat{D}_{kk}}$	0.628	0.669	0.693	0.699	0.000	0.000	0.000	0.000

	$\sqrt{\hat{D}_{2kk}}$ for Survival component[a]							
	Intercept	$Z_{2,1}$	$Z_{2,2}$	$Z_{2,3}$	$Z_{2,4}$	$Z_{2,5}$	$Z_{2,6}$	$Z_{2,7}$
True value $\sqrt{D_{kk}}$	0.707	0.707	0.707	0.707	0	0	0	0
W/O selection $\sqrt{\hat{D}_{kk}}$	1.037	1.074	1.155	1.167	0.574	0.552	0.535	0.685
1^{st} stage $\sqrt{\hat{D}_{kk}}$	0.511	0.431	0.412	0.382	0.002	0.004	0.000	0.025
2^{nd} stage $\sqrt{\hat{D}_{kk}}$	0.652	0.698	0.725	0.782	0.047	0.018	0.005	0.091

[a] $\sqrt{\hat{D}_{1kk}}$ and $\sqrt{\hat{D}_{2kk}}$ are the averages of estimates over the 100 data sets.



Web Figure 1: Residual plots for data application diagnostics. The circles are the standardized residuals. The black lines are the LOESS estimates.