Research Article

Statistics in Medicine

Received XXXX

(www.interscience.wiley.com) DOI: 10.1002/sim.0000

Improving estimation and prediction in linear regression incorporating external information from an established reduced model supplementary materials

Wenting Cheng[a]*, Jeremy M. G. Taylor[a], Pantel S. Vokonas[b,c], Sung Kyun Park[d], Bhramar Mukherjee[a]

1. Web Appendix A

Bootstrap estimate of the standard error for the constrained ML estimate

We would like to obtain a bootstrap estimate of the constrained ML estimator's standard error. Regression models can be bootstrapped by (1) treating the design matrix as random and selecting bootstrap samples directly from the observations or (2) treating the design matrix as fixed and resampling from the residuals of the fitted regression models.[1] In our study, we implement a residual bootstrap as follows:

- Estimate the regression coefficients $\gamma_0, \ldots, \gamma_{p+1}$ and $\theta_0, \ldots, \theta_p$ by constrained ML method for the original sample. Estimate $\boldsymbol{\beta}$ by $\hat{\beta}_j = \hat{\gamma}_j + \hat{\gamma}_{p+1}\hat{\theta}_j, j = 0, \ldots, p$
- Calculate the fitted outcome pair $(\hat{Y}_i, \hat{B}_i)$ and residual pair $\mathbf{E_i} = (E_{i,Y}, E_{i,B})$ for each observation: $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \cdots + \hat{\beta}_p X_{ip}, \hat{B}_i = \hat{\theta}_0 + \hat{\theta}_1 X_{i1} + \cdots + \hat{\theta}_p X_{ip}$ and $E_{i,Y} = Y_i - \hat{Y}_i, E_{i,B} = B_i - \hat{B}_i$
- Take bootstrap samples of the residual pairs, $\tilde{\mathbf{e}}_\mathbf{b} = [\tilde{\mathbf{E}}_\mathbf{b1}, \ldots, \tilde{\mathbf{E}}_\mathbf{bn}]^T, b = 1, \ldots, S$, calculate bootstrapped $\mathbf{Y}$ values $\tilde{\mathbf{Y}}_\mathbf{b} = [\tilde{Y}_{b1}, \ldots, \tilde{Y}_{bn}]^T$, where $\tilde{Y}_{bi} = \hat{Y}_i + \tilde{E}_{bi,Y}$, calculate bootstrapped $\mathbf{B}$ values $\tilde{\mathbf{B}}_\mathbf{b} = [\tilde{B}_{b1}, \ldots, \tilde{B}_{bn}]^T$, where $\tilde{B}_{bi} = \hat{B}_i + \tilde{E}_{bi,B}$
- Regress $\tilde{\mathbf{Y}}_\mathbf{b}$ on the fixed $\mathbf{X}$ design matrix and bootstrap samples $\tilde{\mathbf{B}}_\mathbf{b}$ to obtain bootstrap estimates of regression coefficients by constrained ML: $\tilde{\gamma}_{b,0}, \ldots, \tilde{\gamma}_{b,p+1}$

[a]*Department of Biostatistics, University of Michigan, Ann Arbor, Michigan, USA*
[b] *VA Normative Aging Study, Veterans Affairs Boston Health Care System, Boston, Massachusetts, USA*
[c]*Department of Medicine, Boston University School of Medicine, Boston, Massachusetts, USA*
[d]*Department of Epidemiology, University of Michigan, Ann Arbor, Michigan, USA*
**Correspondence to: Wenting Cheng, Department of Biostatistics, University of Michigan, Ann Arbor, Michigan, USA. Email: chengwt@umich.edu*

- The $\tilde{\gamma}_b$ can be used to construct bootstrap standard error: $\tilde{\sigma}_{.,j} = (\frac{\sum_{b=1}^{S}(\hat{\gamma}_{b,j}-\bar{\tilde{\gamma}}_{.,j})^2}{S-1})^{1/2}$, $j = 0, \ldots, p+1$, in the usual bootstrap manner as described in.[2]

For the partial regression estimate, the bootstrap estimate of the standard error can be obtained in a way similar to that described above.

2. Web Appendix B

Variance approximation for the partial regression estimate

To derive the variance of the partial regression estimate $(\hat{\gamma}_0, \ldots, \hat{\gamma}_{p+1})$, we adapt the variance approximation used in.[3] Since the partial regression solution only uses the point estimates of $\boldsymbol{\beta}$ but not their standard errors as the external information, we treat $\bar{\boldsymbol{\beta}}$ as constants when deriving the variances of $\hat{\boldsymbol{\gamma}}$ and then the partial regression estimates are functions of $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\alpha}}$ only. We can first derive the variance-covariance matrix of $\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\alpha}}$ and then apply delta method to produce the variances of partial regression estimate.

Let $\mathbf{I}^{\theta}{}_{(\mathrm{p+1})\times(\mathrm{p+1})}$ and $\mathbf{I}^{\alpha}{}_{2\times 2}$ denote the observed information matrices for linear regression model $E(\mathbf{B}|\mathbf{X}) = \mathbf{X}\boldsymbol{\theta}$ and for model $\mathrm{E}(\mathbf{r_1}|\mathbf{r_2}) = \alpha_0 + \alpha_1\mathbf{r_2}$ respectively in the partial regression method. $\mathbf{I}^{\theta}{}_{(\mathrm{p+1})\times(\mathrm{p+1})} = \mathbf{X}^{\mathrm{T}}\mathbf{X}$ and $\mathbf{I}^{\alpha}{}_{(2)\times(2)} = \mathbf{V}^{\mathrm{T}}\mathbf{V}$ where $\mathbf{V} = (\mathbf{1}_{(\mathrm{n})\times(1)}, \mathbf{r_2}_{(\mathrm{n})\times(1)})$. Denote the true parameter values of $\boldsymbol{\theta}, \boldsymbol{\alpha}$ by $\boldsymbol{\theta_0}$ and $\boldsymbol{\alpha_0}$. Then asymptotically, $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta_0} = \sqrt{n}(\mathbf{I}^{\theta})^{-1}\sum_{\mathrm{i=1}}^{\mathrm{n}}\mathbf{U}_{\mathbf{i}}^{\theta} + \mathrm{o_p}(\mathrm{n}^{-1/2})$ and $\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha_0} = \sqrt{n}(\mathbf{I}^{\alpha})^{-1}\sum_{\mathrm{i=1}}^{\mathrm{n}}\mathbf{U}_{\mathbf{i}}^{\alpha} + \mathrm{o_p}(\mathrm{n}^{-1/2})$ where $\mathbf{U}_{\mathbf{i}}^{\theta} = (\mathrm{Y_i} - \mathbf{X_i}\boldsymbol{\theta})\mathbf{X_i}$ and $\mathbf{U}_{\mathbf{i}}^{\alpha} = (\mathrm{r_{i1}} - \mathbf{V_i}\boldsymbol{\alpha})\mathbf{V_i}$ are subject i'th individual score functions for estimate $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\alpha}}$ respectively.

The asymptotic variance-covariance matrix of the vector $(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\alpha}})^T$ can be represented as

$$\boldsymbol{\Sigma}_{\boldsymbol{\theta},\boldsymbol{\alpha}} = \begin{pmatrix} \hat{\boldsymbol{\Sigma}}_{\boldsymbol{\theta}} & (\mathbf{I}^{\theta})^{-1}\mathrm{Cov}(\sum_{\mathrm{i=1}}^{\mathrm{n}}\mathbf{U}_{\mathbf{i}}^{\theta}, \sum_{\mathrm{i=1}}^{\mathrm{n}}\mathbf{U}_{\mathbf{i}}^{\alpha})(\mathbf{I}^{\alpha})^{-1\mathrm{T}} \\ (\mathbf{I}^{\alpha})^{-1}\mathrm{Cov}(\sum_{\mathrm{i=1}}^{\mathrm{n}}\mathbf{U}_{\mathbf{i}}^{\alpha}, \sum_{\mathrm{i=1}}^{\mathrm{n}}\mathbf{U}_{\mathbf{i}}^{\theta})(\mathbf{I}^{\theta})^{-1\mathrm{T}} & \hat{\boldsymbol{\Sigma}}_{\boldsymbol{\alpha}} \end{pmatrix}$$

where $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\alpha}}$ are the estimated covariance matrices by OLS and $\mathrm{Cov}(\sum_{\mathrm{i=1}}^{\mathrm{n}}\mathbf{U}_{\mathbf{i}}^{\alpha}, \sum_{\mathrm{i=1}}^{\mathrm{n}}\mathbf{U}_{\mathbf{i}}^{\theta}) = \sum_{\mathrm{i=1}}^{\mathrm{n}}\mathbf{U}_{\mathbf{i}}^{\alpha}\mathbf{U}_{\mathbf{i}}^{\theta\mathbf{T}}$.

Since $\hat{\gamma}_0 = \hat{\alpha}_0 - \hat{\alpha}_1\hat{\theta}_0 + \bar{\beta}_0 = g(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\alpha}})$, by delta method, we have the approximate variance of $\hat{\gamma}_0$: $g'^T\boldsymbol{\Sigma}_{\boldsymbol{\theta},\boldsymbol{\alpha}}g'$. For other γs, we can obtain the approximate variance in a similar way.

We run a simulation study to evaluate the accuracy of the above two variance estimation methods for the regression coefficients for the three-covariate scenario and for the five-covariate scenario. The results are shown in Table 1 and Table 2.

3. Web Appendix C

Posterior distributions in informative full Bayes methods

The conditional distribution of $\beta_0, \ldots, \beta_p$ will be normal, each with distribution function $N(\mu_{\beta_j,n}, \sigma^2_{\beta_j,n}), j = 0, \ldots, p$, where $\mu_{\beta_j,n} = \frac{\sum_{i=1}^{n}(B_i - \sum_{k\neq j}\frac{\beta_k-\gamma_k}{\gamma_{p+1}}X_{ik} + \frac{\gamma_j}{\gamma_{p+1}}X_{ij})X_{ij}\frac{\bar{S}_j^2}{\gamma_{p+1}} + \bar{\beta}_j\sigma_2^2}{\frac{\sum_{i=1}^{n}X_{ij}^2\bar{S}_j^2}{\gamma_{p+1}^2} + \sigma_2^2}$, $\sigma^2_{\beta_j,n} = \frac{\sigma_2^2\bar{S}_j^2}{\frac{\sum_{i=1}^{n}X_{ij}^2\bar{S}_j^2}{\gamma_{p+1}^2} + \sigma_2^2}$.

The conditional distribution of $\gamma_0, \ldots, \gamma_p$ will be normal, each with distribution function $N(\mu_{\gamma_j,n}, \sigma^2_{\gamma_j,n}), j = 0, \ldots, p$, where $\mu_{\gamma_j,n} = \frac{[\sum_{i=1}^{n}(Y_i - \sum_{k\neq j}\gamma_k X_{ik} - \gamma_{p+1}B_i)X_{ij}\sigma_2^2 - \sum_{i=1}^{n}(B_i - \sum_{k\neq j}\frac{\beta_k-\gamma_k}{\gamma_{p+1}}X_{ik} - \frac{\beta_j X_{ij}}{\gamma_{p+1}})X_{ij}\frac{\sigma_1^2}{\gamma_{p+1}}]\times 100^2}{(\sum_{i=1}^{n}X_{ij}^2\sigma_2^2 + \frac{\sum_{i=1}^{n}x_{ij}^2}{\gamma_{p+1}^2}\sigma_1^2)100^2 + \sigma_1^2\sigma_2^2}$, and $\sigma^2_{\gamma_j,n} = \frac{\sigma_1^2\sigma_2^2 100^2}{(\sum_{i=1}^n X_{ij}^2\sigma_2^2 + \frac{\sum_{i=1}^n X_{i2}^2}{\gamma_{p+1}^2}\sigma_1^2)100^2 + \sigma_1^2\sigma_2^2}$.

The conditional distribution of σ_1^2, σ_2^2 will be inverse-gamma. The full conditional distribution for σ_1^2 is $\text{inverse}-\text{gamma}(\frac{\nu_{1,\text{n}}}{2}, \frac{\nu_{1,\text{n}}\sigma^2_{1,\text{n}}}{2})$ where $\nu_{1,n} = \nu_0 + n$ and $\sigma^2_{1,n} = \frac{1}{\nu_{1,n}}[\sum_{i=1}^n (Y_i - \sum_{j=0}^p \gamma_j X_{ij} - \gamma_{p+1}B_i)^2 + \nu_0\sigma_0^2]$. The full conditional distribution for σ_2^2 is $\text{inverse}-\text{gamma}(\frac{\nu_{2,\text{n}}}{2}, \frac{\nu_{2,\text{n}}\sigma^2_{2,\text{n}}}{2})$ where $\nu_{2,n} = \nu_0 + n$ and $\sigma^2_{2,n} = \frac{1}{\nu_{2,n}}[\sum_{i=1}^n (B_i - \sum_{j=0}^p \frac{\beta_j - \gamma_j}{\gamma_{p+1}} X_{ij})^2 + \nu_0\sigma_0^2]$.

The full conditional distribution of γ_{p+1} is: $p(\gamma_{p+1}|\mathbf{Y}, \mathbf{X}, \mathbf{B}, \beta_0, \ldots, \beta_p, \gamma_0, ..., \gamma_p, \sigma_1^2, \sigma_2^2)$ $\propto \left\{\prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{1}{2\sigma_1^2}(Y_i - \sum_{j=0}^p \gamma_j X_{ij} - \gamma_{p+1}B_i)^2} \times \frac{1}{\sqrt{2\pi\sigma_2^2}} e^{-\frac{1}{2\sigma_2^2}(B_i - \sum_{j=0}^p \frac{\beta_j - \gamma_j}{\gamma_{p+1}} X_{ij})^2}\right\} \cdot \pi(\gamma_{p+1}) \cdot |\mathbf{J}|$, which does not have a closed form.

4. Web Appendix D

Characteristics of the training dataset and the characteristics of the testing dataset

The characteristics of the training dataset and the characteristics of the testing dataset are shown in Table 3.

5. Web Appendix E

Sensitivity analysis of the choice of d for the constrained ML in the expanded tibia lead prediction model

In the nonlinear constraints, d plays an important role. d represents our belief in the strength of the external/historical information. In the manuscript we suggested that fix d as d = 1. The choice $d = 0$ assumes that there is no uncertainty in the regression coefficient estimates in the established historical data model. We evaluate the sensitivity of the choice of d in the bone lead data example. The results are summarized in Table 4. It shows that the choice of d will influence both the estimation efficiency of the regression coefficients and the predictive power of the model. With bigger d, the standard errors of the regression coefficients will increase, the predictive power in the validation dataset will decrease. When $d = 10$, these box constraints are observed to be weak and the estimated model based on the constrained methods are more similar to the estimated model based on direct regression on just the internal data. The small differences between the standard errors for $d = 10$ case and those from direct regression are due to the differences in the bootstrap procedure applied when $d = 10$ and information matrix based standard errors for the direct regression method.

6. Web Appendix F

Additional simulations with different sample sizes

We conducted additional simulations for the case that n = 50; 100; 200; 2000 for the three-covariate simulation scenario and the five-covariate simulation scenario. The results are summarized in Table 5 and Table 6.

References

1. Fox J. *Applied regression analysis and generalized linear models*. Sage, 2008.
2. Efron B, Tibshirani R. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science* 1986; **1**(1):54–75.
3. Mukherjee B, Chatterjee N. Exploiting gene-environment independence for analysis of case-control studies: an empirical bayes-type shrinkage estimator to trade-off between bias and efficiency. *Biometrics* 2008; **64**(3):685–694. URL `http://www.jstor.org/stable/25502124`.

Table 1. Simulation results of the three-covariate scenario: for both the constrained ML and the partial regression, we report the ratio of average bootstrap mean and Monte Carlo mean $((\frac{1}{500}\sum_{m=1}^{500}\bar{\bar{\gamma}}_{m,j})/(\frac{1}{500}\sum_{\mathrm{m=1}}^{500}\hat{\gamma}_{\mathrm{m,j}}))$ and the ratio of average bootstrap standard error and Monte Carlo standard deviation $((\frac{1}{500}\sum_{m=1}^{500}\tilde{\sigma}_{m,j})/\sqrt{V(\hat{\gamma}_j)})$ of each regression coefficient. For the partial regression solution, we also report the ratio of average asymptotic standard error and Monte Carlo standard deviation $(\frac{1}{500}\sum_{\mathrm{m=1}}^{500}\mathrm{Asy.SE}(\gamma_{\mathrm{m,j}})/\sqrt{V(\hat{\gamma}_j)})$

Sample Size	Method	Ratio	γ_1	γ_2	γ_3
$n = 15$	Constrained ML	Avg.boot.Mean/MC.Mean	0.93	0.92	1.12
		Avg.Boot.SE/MC.SD	1.07	1.41	1.97
	Partial regression	Avg.boot.Mean/MC.Mean	1.00	1.00	1.01
		Avg.Boot.SE/MC.SD	1.14	1.15	1.10
		Avg.Asy.SE/MC.SD	1.05	1.04	0.96
$n = 30$	Constrained ML	Avg.boot.Mean/MC.Mean	0.96	0.96	1.06
		Avg.Boot.SE/MC.SD	0.95	1.03	1.09
	Partial regression	Avg.boot.Mean/MC.Mean	1.00	1.00	1.00
		Avg.Boot.SE/MC.SD	0.94	1.08	1.03
		Avg.Asy.SE/MC.SD	0.91	1.03	1.00
$n = 200$	Constrained ML	Avg.boot.Mean/MC.Mean	1.00	1.00	1.00
		Avg.Boot.SE/MC.SD	1.00	1.00	1.03
	Partial regression	Avg.boot.Mean/MC.Mean	1.00	1.00	1.00
		Avg.Boot.SE/MC.SD	1.00	1.06	1.03
		Avg.Asy.SE/MC.SD	1.00	1.06	1.03

Table 2. Simulation results of the five-covariate scenario: for both the constrained ML and the partial regression, we report the ratio of average bootstrap mean and Monte Carlo mean $((\frac{1}{500}\sum_{m=1}^{500}\bar{\bar{\gamma}}_{m,j})/(\frac{1}{500}\sum_{\mathrm{m=1}}^{500}\hat{\gamma}_{\mathrm{m,j}}))$ and the ratio of average bootstrap standard error and Monte Carlo standard deviation $((\frac{1}{500}\sum_{m=1}^{500}\tilde{\sigma}_{m,j})/\sqrt{V(\hat{\gamma}_j)})$ of each regression coefficient. For the partial regression solution, we also report the ratio of average asymptotic standard error and Monte Carlo standard deviation $(\frac{1}{500}\sum_{\mathrm{m=1}}^{500}\mathrm{Asy.SE}(\gamma_{\mathrm{m,j}})/\sqrt{V(\hat{\gamma}_j)})$

Sample Size	Method	Ratio	γ_1	γ_2	γ_3	γ_4	γ_5
$n = 20$	Constrained ML	Avg.boot.Mean/MC.Mean	0.89	0.89	1.00	1.00	1.16
		Avg.Boot.SE/MC.SD	1.13	0.99	1.06	0.97	1.40
	Partial regression	Avg.boot.Mean/MC.Mean	1.00	1.00	1.00	1.00	1.00
		Avg.Boot.SE/MC.SD	1.15	1.03	1.02	0.98	1.17
		Avg.Asy.SE/MC.SD	1.06	0.94	0.97	0.94	1.01
$n = 30$	Constrained ML	Avg.boot.Mean/MC.Mean	0.93	0.93	1.00	1.00	1.11
		Avg.Boot.SE/MC.SD	1.08	1.00	0.94	1.07	1.10
	Partial regression	Avg.boot.Mean/MC.Mean	1.00	1.00	1.00	1.00	1.00
		Avg.Boot.SE/MC.SD	1.03	1.01	0.94	1.05	1.05
		Avg.Asy.SE/MC.SD	0.99	0.97	0.96	1.05	0.95
$n = 200$	Constrained ML	Avg.boot.Mean/MC.Mean	0.99	0.99	1.01	0.99	1.01
		Avg.Boot.SE/MC.SD	1.03	1.00	1.00	1.00	1.00
	Partial regression	Avg.boot.Mean/MC.Mean	1.00	1.00	1.00	1.00	1.00
		Avg.Boot.SE/MC.SD	1.03	1.00	1.00	1.07	1.00
		Avg.Asy.SE/MC.SD	1.03	1.00	1.00	1.07	1.00

Table 3. Characteristics and lead biomarkers of subjects in training dataset (N = 100) and in testing dataset (N = 56)

	Training dataset		Testing dataset	
	Mean ± SD	Range	Mean ± SD	Range
Age (yr)	67.22 ± 7.44	51.64 - 92.25	68.72 ± 6.65	53.93 - 82.42
Cumulative cigarette (pack-yr)	18.32 ± 24.45	0.00 - 136.00	21.50 ± 26.93	0.00 - 105.50
Genetic risk score	11.45 ± 2.27	5.76 - 16.97	10.85 ± 2.36	4.00 - 16.00
Lead biomarkers				
Blood lead (μg/dL)	5.53 ± 3.29	0.00 - 17.00	5.73 ± 2.29	1.00 - 11.00
Tibia lead (μg/g)	20.59 ± 11.86	3.00 - 76.00	20.89 ± 14.06	-1.00 - 77.00
		N(%)		
Smoking status				
Never		39 (39.00)		16 (28.57)
Former		52 (52.00)		38 (67.86)
Current		9 (9.00)		2 (3.57)
Education				
High school dropout		10 (10.00)		10 (17.86)
High school diploma		58 (58.00)		30 (53.57)
≥ 4 yr of college		32 (32.00)		16 (28.57)
White collar		61 (61.00)		31 (55.36)

Table 4. Regression coefficients of the expanded tibia lead prediction model with the genetic score (n = 100)

Variable	Constrained ML				Direct Regression
	$d = 0$	$d = 1$	$d = 5$	$d = 10$	
Intercept	-21.90(6.56)	-27.47(7.27)	-26.94(10.57)	-23.85(10.72)	-23.85(11.27)
Blood lead	1.05(0.08)	0.93(0.13)	0.91(0.27)	0.95(0.29)	0.95(0.30)
Age	0.59(0.02)	0.64(0.06)	0.60(0.12)	0.55(0.13)	0.55(0.13)
Education					
High school diploma	-3.67(0.44)	-1.92(1.42)	2.34(2.79)	3.19(3.20)	3.19(3.44)
≥ 4 yr of college	-7.12(0.60)	-5.04(1.60)	-0.80(3.41)	-0.06(3.82)	-0.06(4.08)
White collar	-3.24(0.32)	-2.84(1.02)	-3.98(2.29)	-4.48(2.29)	-4.48(2.42)
Cumulative cigarette smoking	0.03(0.01)	0.06(0.02)	0.16(0.04)	0.22(0.05)	0.22(0.06)
Smoking status					
Former smoker	1.83(0.35)	2.89(0.98)	0.05(2.05)	-1.92(2.22)	-1.92(2.54)
Current smoker	0.18(0.85)	-2.28(1.90)	-12.21(3.30)	-19.07(4.59)	-19.07(5.48)
Genetic risk score	0.13(0.51)	0.15(0.48)	0.13(0.44)	0.13(0.43)	0.13(0.44)
R^2	0.32	0.35	0.41	0.42	0.42
OOB R^2	0.29	0.29	0.23	0.17	0.17

Table 5. Simulation results of three-covariate scenario: Comparison of different methods. OOB R^2 denotes average out-of-bag prediction ability. For each method, each row includes mean (Monte Carlo standard error) of each regression coefficient and OOB R^2 of this method. A linear regression on $\mathbf{Y}$ on $\mathbf{X_1}, \mathbf{X_2}$ has an OOB R^2 of 0.212

Method	Sample size	γ_1	γ_2	γ_3	OOB R^2
True value		3	3	2	
Direct regression	$n = 15$	3.25(2.27)	3.07(2.43)	1.96(1.39)	0.270
	$n = 50$	2.95(1.08)	2.98(1.07)	2.01(0.58)	0.429
	$n = 100$	2.96(0.75)	3.01(0.78)	2.01(0.44)	0.451
	$n = 200$	2.98(0.52)	2.97(0.54)	2.01(0.30)	0.464
	$n = 2000$	3.01(0.16)	3.00(0.17)	2.00(0.09)	0.474
Constrained ML	$n = 15$	2.82(1.80)	2.79(1.64)	2.27(1.55)	0.334
	$n = 50$	2.92(0.69)	2.93(0.66)	2.08(0.59)	0.453
	$n = 100$	2.95(0.46)	2.98(0.47)	2.05(0.44)	0.462
	$n = 200$	2.97(0.34)	2.98(0.34)	2.03(0.30)	0.469
	$n = 2000$	3.01(0.13)	3.00(0.14)	2.00(0.09)	0.474
Partial regression	$n = 15$	3.03(1.51)	3.01(1.50)	1.96(1.39)	0.346
	$n = 50$	2.99(0.70)	2.99(0.65)	2.01(0.58)	0.446
	$n = 100$	3.00(0.45)	3.01(0.45)	2.01(0.44)	0.459
	$n = 200$	2.99(0.33)	3.00(0.31)	2.01(0.30)	0.468
	$n = 2000$	3.00(0.10)	3.00(0.09)	2.00(0.09)	0.474
Standard Bayes	$n = 15$	3.24(2.28)	3.06(2.44)	1.97(1.39)	0.270
	$n = 50$	2.95(1.08)	2.98(1.07)	2.01(0.58)	0.429
	$n = 100$	2.96(0.75)	3.01(0.78)	2.01(0.44)	0.451
	$n = 200$	2.98(0.52)	2.97(0.54)	2.01(0.30)	0.464
	$n = 2000$	3.01(0.16)	3.00(0.17)	2.00(0.09)	0.474
Informative full Bayes	$n = 15$	3.06(1.42)	2.99(1.43)	1.98(1.33)	0.382
	$n = 50$	2.98(0.65)	2.99(0.62)	2.01(0.56)	0.457
	$n = 100$	2.99(0.44)	3.01(0.44)	2.01(0.43)	0.464
	$n = 200$	2.99(0.32)	2.99(0.30)	2.01(0.30)	0.470
	$n = 2000$	3.01(0.11)	3.00(0.12)	2.00(0.09)	0.475
Transformation	$n = 15$	3.16(1.55)	3.09(1.60)	1.84(1.49)	0.366
	$n = 50$	2.98(0.69)	2.99(0.65)	2.00(0.61)	0.455
	$n = 100$	2.99(0.46)	3.00(0.48)	2.01(0.48)	0.463
	$n = 200$	2.98(0.33)	2.99(0.33)	2.01(0.32)	0.469
	$n = 2000$	3.01(0.11)	3.00(0.12)	2.00(0.10)	0.475

Table 6. Simulation results of five-covariate scenario: Comparison of different methods. OOB R^2 denotes average out-of-bag prediction ability. For each method, each row includes mean (Monte Carlo standard error) of each regression coefficient and OOB R^2 of this method. A linear regression on $\mathbf{Y}$ on $\mathbf{X_1}, \mathbf{X_2}, \mathbf{X_3}, \mathbf{X_4}$ has an OOB R^2 of 0.350

Method	Sample size	γ_1	γ_2	γ_3	γ_4	γ_5	OOB R^2
True value		3	3	2	2	2	
Direct regression	$n = 20$	3.11(1.88)	3.02(2.07)	1.93(1.15)	2.04(1.15)	1.92(1.10)	0.421
	$n = 50$	2.97(1.02)	3.02(1.03)	1.95(0.61)	2.01(0.59)	1.99(0.57)	0.545
	$n = 100$	3.04(0.75)	3.05(0.69)	2.00(0.42)	2.03(0.44)	1.99(0.43)	0.569
	$n = 200$	2.94(0.48)	2.95(0.48)	2.00(0.30)	2.00(0.29)	2.04(0.29)	0.581
	$n = 2000$	3.00(0.15)	3.00(0.16)	2.00(0.09)	2.00(0.09)	1.99(0.09)	0.591
Constrained ML	$n = 20$	2.77(1.36)	2.66(1.60)	2.03(0.70)	1.94(0.75)	2.33(1.26)	0.492
	$n = 50$	2.89(0.67)	2.93(0.67)	2.05(0.30)	1.94(0.31)	2.12(0.59)	0.567
	$n = 100$	2.97(0.49)	2.99(0.46)	2.07(0.21)	1.95(0.22)	2.05(0.43)	0.577
	$n = 200$	2.95(0.31)	2.93(0.33)	2.06(0.15)	1.94(0.15)	2.07(0.29)	0.582
	$n = 2000$	3.00(0.13)	3.00(0.13)	2.03(0.06)	1.97(0.06)	2.00(0.09)	0.589
Partial regression	$n = 20$	3.08(1.21)	3.01(1.37)	2.09(0.64)	1.91(0.65)	1.92(1.10)	0.500
	$n = 50$	3.00(0.65)	3.03(0.67)	2.10(0.33)	1.91(0.31)	1.99(0.57)	0.560
	$n = 100$	3.01(0.48)	3.03(0.46)	2.10(0.21)	1.91(0.22)	1.99(0.43)	0.572
	$n = 200$	2.99(0.31)	2.96(0.32)	2.10(0.15)	1.90(0.14)	2.04(0.29)	0.577
	$n = 2000$	3.00(0.10)	3.01(0.10)	2.10(0.04)	1.90(0.04)	1.99(0.09)	0.583
Standard Bayes	$n = 20$	3.11(1.88)	3.02(2.07)	1.93(1.15)	2.04(1.15)	1.92(1.10)	0.421
	$n = 50$	2.97(1.02)	3.02(1.03)	1.95(0.61)	2.01(0.59)	1.99(0.58)	0.545
	$n = 100$	3.04(0.75)	3.05(0.70)	2.00(0.42)	2.03(0.44)	1.99(0.43)	0.569
	$n = 200$	2.94(0.48)	2.95(0.48)	2.00(0.30)	2.00(0.29)	2.04(0.29)	0.580
	$n = 2000$	3.00(0.15)	3.00(0.16)	2.00(0.09)	2.00(0.09)	1.99(0.09)	0.591
Informative full Bayes	$n = 20$	3.02(1.09)	2.94(1.30)	2.06(0.59)	1.92(0.62)	1.99(0.99)	0.526
	$n = 50$	2.99(0.60)	3.03(0.61)	2.06(0.28)	1.93(0.28)	1.99(0.54)	0.570
	$n = 100$	3.02(0.46)	3.03(0.43)	2.08(0.19)	1.93(0.20)	1.99(0.42)	0.578
	$n = 200$	2.97(0.30)	2.96(0.31)	2.07(0.13)	1.93(0.13)	2.04(0.29)	0.582
	$n = 2000$	3.00(0.11)	3.00(0.11)	2.04(0.05)	1.96(0.06)	1.99(0.09)	0.588
Transformation	$n = 20$	3.16(1.23)	3.08(1.44)	2.04(0.62)	1.95(0.64)	1.81(1.22)	0.516
	$n = 50$	3.01(0.65)	3.05(0.67)	2.06(0.29)	1.93(0.29)	1.96(0.64)	0.568
	$n = 100$	3.03(0.49)	3.04(0.46)	2.07(0.20)	1.94(0.20)	1.97(0.46)	0.578
	$n = 200$	2.97(0.31)	2.95(0.32)	2.07(0.14)	1.93(0.13)	2.05(0.30)	0.582
	$n = 2000$	3.00(0.11)	3.00(0.11)	2.05(0.05)	1.95(0.05)	2.00(0.09)	0.588

